

①9 BUNDESREPUBLIK

DEUTSCHLAND



DEUTSCHES  
PATENT- UND  
MARKENAMT

⑫

Offenlegungsschrift

⑩

DE 100 04 930 A 1

⑤①

Int. Cl. 7:

C 07 K 16/00

C 12 N 15/63

C 12 N 15/11

- ②① Aktenzeichen: 100 04 930.3  
②② Anmeldetag: 4. 2. 2000  
④③ Offenlegungstag: 9. 8. 2001

DE 100 04 930 A 1

⑦① Anmelder:

Brüß, Michael, Dr., 53121 Bonn, DE; Bönisch, Heinz,  
Prof. Dr., 53125 Bonn, DE

⑦② Erfinder:

gleich Anmelder

Die folgenden Angaben sind den vom Anmelder eingereichten Unterlagen entnommen

⑤④ Seralpha: Gen, cDNA, Expression, Aminosäuresequenz

DE 100 04 930 A 1

## Beschreibung

Mittels Homologiesuche in Datenbanken wurde ein Teil eines potentiellen neuen Gens welches für einen G-Protein-gekoppelten Rezeptor kodiert identifiziert. Aus der Partialsequenz wurden Oligonukleotid-Primer abgeleitet und für eine PCR Amplifikation eingesetzt. Mittels dieser Primer und der Polymerasekettenreaktion konnten wir das Rezeptorgen aus genomischer humaner DNA amplifizieren und mit dieser Methode konnten wir auch die cDNA welche für diesen Rezeptor kodiert aus humaner Gehirn cDNA amplifizieren. Es handelt sich um ein Intron-lostes Gen, d. h. die genomische Sequenz und die cDNA sind identisch. Ferner konnten wir auf genomischer DNA zusätzlich 5' und 3' nicht translatierte Bereiche des Gens (mittels genome walking) identifizieren und sequenzieren. Dargestellt und beschrieben werden das humane Gen für den von uns als SERALPHA bezeichneten neuen Neurotransmitter Rezeptor, die cDNA welche den Rezeptor kodiert und die von der cDNA abgeleitete Aminosäuresequenz. Aufgrund der cDNA Amplifikation aus humaner Gehirn cDNA wissen wir ferner, daß dieser neue, bisher unbekannte Rezeptor im menschlichen Gehirn exprimiert wird.

1. Das Gen inclusive 5' und 3' nichttranslatiertem Bereich welches wir zur Patentierung anmelden, ist 4260 Basen groß. Der kodierende Bereich erstreckt sich von Base 1317 (Start Codon ATG) bis Base 2843 (Stop Codon TGA inclusive). Im nicht translatierten 5' Bereich (Base 1–1316) befindet sich der regulative Bereich dieses Gens (Promotor) eine TATA Box befindet sich im Bereich von 699–713. Im nichttranslatierten 3' Bereich des Gens (Base 2841 bis 4260) befindet sich ein klassisches Polyadenylierungssignal bei Position 4169–4174 (AATAAA), außerdem findet man mehrere putative Polyadenylierungssignale schon weiter 3' gelegen.

2. Die cDNA welche wir aus humaner Gehirn cDNA amplifizieren konnten erstreckt sich von Position 1300 bis 2863 des Gens und kodiert für ein Protein mit 508 Aminosäuren. Für die Amplifikation wurden folgende Primer benutzt: sense: 5'GCCCTCTGGCTGTTGCCATG3'; antisense: 5'GGTTTGCCTTAGAACTAACTTC3'. Durch die Klonierung der cDNA aus humanem Gehirn ist bewiesen, daß dieser Rezeptor existiert und in menschlichem Gehirn exprimiert wird.

3. Die Aminosäuresequenz welche aus dem offenen Leseraster der cDNA und auch des intronlosen Gens abgeleitet werden kann ist 508 Aminosäuren lang. Homologievergleiche in Datenbanken ergibt, daß das Protein die höchste Homologie zu Neurotransmitter Rezeptoren aufweist. Die höchsten Homologien wurden zum humanen Serotonin 1A Rezeptor, zum muskarinischen Acetylcholin Rezeptoren (M4) und zu adrenergen Rezeptoren (alpha 1A und 1C) gefunden. Hydrophobizitätsanalyse der Aminosäuresequenz mit dem Programm TOPPED2 ergab, daß das Protein die typische Struktur von G-Protein gekoppelten Rezeptoren mit sieben transmembranalen Domänen besitzt. Ausserdem wurde in der Sequenz ein typischen Motiv für G-Protein gekoppelte Rezeptoren (Aminosäure 117–133: ASVNTIVVVSVDRLSI) gefunden. Im N-terminalen Bereich des Rezeptor der vermutlich (wie bei allen G-Protein gekoppelten Rezeptoren) extrazellulär liegt befinden sich zwei potentielle N-Glykosylierungsstellen (Pos. 7 und 13). Zwischen transmembranaler Domäne 5 und 6 befindet sich eine große intrazelluläre Schleife aus fast 200 Aminosäuren. Potentielle Phosphorylierungsstellen für Protein-Kinase C befinden sich in den intrazellulären Positionen: 66; 144; 283; 324; 331; 379; 460. An den Positionen 233 und 283 kann potentiell durch cAMP und cGMP-abhängige Protein-Kinasen phosphoryliert werden.

4. Einordnung und potentielle Funktionen des zu patentierenden Rezeptors und seines Gens (bzw cDNA; mRNA)

Der Rezeptor gehört höchstwahrscheinlich zur großen Genfamilie der G-Protein-gekoppelten Rezeptoren (GPCRs). Innerhalb dieser Großfamilie zählt er zur Sub-Familie der "Klasse A Rhodopsin-ähnlichen" Rezeptoren. Er besitzt hohe Homologie zu Rezeptoren für Neurotransmitter (besonders für Serotonin, Adrenalin/Noradrenalin, Acetylcholin und Dopamin). Aufgrund dieser Homologien und weil der Rezeptor im menschlichen Gehirn exprimiert wird, handelt es sich höchstwahrscheinlich um einen wichtigen, bisher unbekannten neuen Neurotransmitter-Rezeptor. Neurotransmitter-Rezeptoren vermitteln Signale nachdem ein Neurotransmitter an einer chemischen Synapse ausgeschüttet wurde. Der entsprechende Neurotransmitter kann dann an der postsynaptischen Membran über Rezeptoren das Signal weiterleiten (second messenger Kaskade) und er kann auf der präsynaptischen Membran über Autorezeptoren seine eigene Freisetzung modulieren. Fast alle bekannten Neurotransmitter Rezeptoren stellen wichtige Angriffspunkte für therapeutisch bedeutsame Pharmaka dar. Das trifft auch für den von uns neu gefundenen Rezeptor potentiell zu.

## MBHB-Rezept-7: SERALPHA: Gensequenz

	10	20	30	40	50	
1	GAAAAACCAAGGTGAGGGGCCATGGGAAAAAGATGTAGAGAAATAGAGAT					5
51	GCGCACACACTCAAAGATACAAAGCAAGTGTTCAGCCAGAGTCCTGGGGT					
101	GACAGGAAAGTGAAGGACAGAGAAACCAACGAGAGAGTGCACACACTAAA					10
151	GAGATAGCTTGAGAGACAGGTGAGACAAGCAGTAGAGTGTTAGAGATCAA					
201	TGGAGAGAGAGGCAGAGTAATGCAGAGAGAACTTAGGCAGGGAGAGACAG					
251	AAACCACACGGAGAGACACAGAGACAAGAGGGTCCGAGAAAGACAGAGCA					
301	AGCCGACACGTGCCAGTGAGAGACAGAAGGAGACACCCAGAGAACCCGAG					
351	GGAGACGCAGAAACAAGAGAGTTGGAGAGACACGGAGACGGGCAGAGACA					15
401	CAGCGAACGACATTGAACAGGTGCGCTAGGGACCCACAAAGATGAACTGC					
451	TGGGGAATAGACTGAGAGAGAAATGGAGAGACAAGAGAGCACTAGGGAA					
501	AGACGGAGACACACATGAACTAGCGAGCGACACAGTGACAGGCAGAGACA					
551	CAGAGAGACAGAATGGATCAGAAAGAGACTGAGGGAGAGACGGGCAAGAC					
601	TGTGAGAAAGATAGGGAGCGAGAGCGCAAGCCCTGCAGAGAAAGTGCCT					20
651	AGCGAAAGAGACACACACAGAGAGACTGAGTGACAAACAGAGTGGAAGAG					
701	ATAAATGAGAGTAACAGGAAAGAGACAGCGTGCACGTGCTATCGAGAGAC					
751	GGAGAAACAGAGACAGACATAAGTGAGAGACAAGGATGAGAAAGAATCAG					
801	GAAACGGTGAGACACATACAGACACACAACCCGAGAGATAAAGAAAAAAA					
851	GTGCGCTAGCGAGAGACACTGGAGAAGAGACAGAAACTGACGGTCGTAAA					25
901	CAGACAGAAACAGAGAGCGCACGCGCATGCGCTAGCGCTAGCGACCAAAA					
951	CTCCCCGAGCCAGAGACAGTGTGAGAGACAAGCAGAGAAAGCGCACGCA					
1001	CGCACGCGCCAGCAGGAGAAACAGATGAGAGGAAATCAGAGCCCTGGAGA					
1051	GAGACAGGCAGACAGATCTGGAGAGTCCGAAAGGAGCCATAGAAGCTGC					
1101	CCGCACTGGGGATGGAGCCGTGCGGAAACCCGGGGTAGGGGGTCTTCAG					30
1151	CGTCCTTGCTGGGCGCGGAGGCTTCTCCCTTGACGGGTGACTAATCTG					
1201	CCTGCGTGTCTTTCTTTGTCAACAGCATAGGCACTGAGTGCGGTCTGTGCA					
1251	CCCCTTTGCCACCCAGCGTGCCGGCACTGAGCCTGCAACCTGTCTCACG					
1301	CCCTCTGGCTGTTGCCATGACGTCCACCTGCACCAACAGCAQCGCGAGA					
1351	GTAACAGCAGCCACACGTGCATGCCCTCTCCAAATGCCATCAGCCTG					35
1401	GCCACGGCATCATCCGCTCAACCGTGCTGGTTATCTTCCTCGCCGCCTC					
1451	TTTCGTCGGCAACATAGTGCTGGCGCTAGTGTTGCAGCGCAAGCCGCAGC					
1501	TGCTGCAGGTGACCAACCGTTTTATCTTTAACCTCCTCGTCACCGACCTG					
1551	CTGCAGATTTTCGCTCGTGGCCCCCTGGGTGGTGGCCACCTCTGTGCCTCT					
1601	CTTCTGGCTGTCCTCAACAGCCACTTCTGCACGGCCCTGGTTAGCCTCACC					40
1651	ACCTGTTTCGCTTCGCCAGCGTCAACACCATTGTGCTGGTGTGAGTGGAT					
1701	CGCTACTTGTCCATCATCCACCCTCTCTCCTACCCGTCCAAGATGACCCA					
1751	GCGCCGCGGTTACCTGCTCCTCTATGGCACCTGGATTGTGGCCATCCTGC					
1801	AGAGCACTCCTCCACTCTACGGCTGGGGCCAGGCTGCCTTTGATGAGCGC					
1851	AATGCTCTCTGCTCCATGATCTGGGGGGCCAGCCCCAGCTACACTATTCT					45
1901	CAGCGTGGTGTCCTTCTCATCGTCATTCCACTGATTGTGTCATGATGCTGCT					
1951	ACTCCGTGGTGTCTGTGTCAGCCCGGAGGCAGCATGCTCTGCTGTACAAT					
2001	GTCAAGAGACACAGCTTGGAAGTGCAGTCAAGGACTGTGTGGAGAATGA					
2051	GGATGAAGAGGGAGCAGAGAAGAAGGAGGAGTTCCAGGATGAGAGTGAGT					
2101	TTCCGCCCCAGCATGAAGGTGAGGTCAAGGCCAAGGAGGGCAGAATGGAA					50
2151	GCCAAGGACGGCAGCCTGAAGGCCAAGGAAGGAAGCACGGGGACCAAGTGA					
2201	GAGTAGTGTAAGGCCAGGGGCAGCGAGGAGGTGAGAGAGAGCAGCACGG					
2251	TGGCCAGCGACGGCAGCATGGAGGGTAAGGAAGGCAGCACCAAGTTGAG					
2301	GAGAACAGCATGAAGGCAGACAAGGGTCGCACAGAGGTCAACCAGTGCAG					
2351	CATTGACTTGGGTGAAGATGACATGGAGTTTGGTGAAGACGACATCAATT					55
2401	TCAGTGAGGATGACGTCGAGGCAGTGAACATCCCGGAGAGCCTCCCACCC					
2451	AGTCGTGTAACAGCAACAGCAACCCTCCTCTGCCCAGGTGCTACCAAGTGA					
2501	CAAAGCTGCTAAAGTGATCTTCATCATCATTTTCTCCTATGTGCTATCCC					60
2551	TGGGGCCCTACTGCTTTTTTAGCAGTCTTGGCCGTGTGGGTGGATGTGAA					
2601	ACCCAGGTACCCAGTGGGTGATCACCATAATCATCTGGCTTTTCTTCCT					
2651	GCAGTGCTGCATCCACCCTATGTCTATGGCTACATGCACAAGACCATTGA					
2701	AGAAGGAAATCCAGGACATGCTGAAGAAGTTCTTCTGCAAGGAAAAGCCC					
2751	CCGAAAGAAGATAGCCACCCAGACCTGCCCGGAACAGAGGGTGGGACTGA					
2801	AGGCAAGATTGTCCCTTCCTACGATTCTGCTACTTTTCCTTGAAGTTAGT					65

2851 TCTAAGGCAAACCTTGAAGTGTCCATAACACGAGAAACAAGAGGAGATTT  
 2901 CTTTTCAATGGACCCACAATTCATTAATGCCAAACCATAACATTTTCAGGC  
 2951 AAAGGTGTTGCACACACATGCTCTTCACCACAAGGTAGATAAATATATAG  
 3001 AAGAGGCAGGAACTGGGGTCTTTCCGTAAAAGCATGGACTTGAGGATTCT  
 3051 GACTGAAATTTTCCCCCAAAGATTATTAGGCTCTACATTTCTTAAAGCA  
 3101 ACAAGGGCTATCCATTTTGGACTTGTAGTTGGTATTCTATCTTTTCCAGA  
 3151 GCTACAACATGCCAACTTTAGCTCTGAAGGAAAGGGAAGATGATGCTTGT  
 3201 GAACTTAAGGACTTTTCGGCCCTCGGGTCGGGAGCTCATGGGCCAGAGCT  
 3251 ACAGCTTGTGTTCAACTGAAAGAAAGGCAATGGACCAATCATTTCATGGA  
 3301 GCCCAAGAAACAGAACCTAATGGACTGATCAACATATGAGCCAAATTCTG  
 3351 AACTGAACAGCCCCACAGTCGGGTGCAAAGACTGTTACACAAACTAAAC  
 3401 AAAGGGCTCCTACAGTTAGAATCTCAAGAAGGTTTCTAGATCCCCTAAA  
 3451 GGGATCCAGAAAAGTAGAAGGACATGTATGAAATGGGAAGCTAGTCCAAG  
 3501 GGAAAAGATTGAGAAATAACACACATCTGGAGAGCTAAACAGTTGACTTT  
 3551 TTTTCTATAAAATCTTGGGTTTATGCATGGGCTGGAAGTGAAGTTCATTA  
 3601 AGTGTAATTTGTCAATTGACACAAATATTTTCTGTCTCCTGTTTGAATAA  
 3651 TAGTGGGGCAGAAATCATGCCACTATTTTACAACCTCCCTTATGTGACTG  
 3701 AATTGAGATGCTGGTGGGAATTCTTCAGATCTCTGCCAACACTTCTGTTT  
 3751 TCTTTTGGTTTGTGTTTGTCAAATAAGCCTTTTCTAGTCAAACAGTATTT  
 3801 ACAGAAAAAAGAAATTCAACTAGAAGTGGCCTAAGTCTACAAAATTCA  
 3851 TGATGTCACTGAGGAATAATTTGTTTCATCAGAAATATATTTTGTGTCCAT  
 3901 GAGATCATAGACAATAAATGTGATCTCCACATGGGGAGCAAGGAAGGCAG  
 3951 AATGAACATTTTTCTTCTCCAGGCACACCCATGTGTCTTTTCCACCTGT  
 4001 GGCTCTCTTTAAAGCTTTTAAAGCTCTCTGCAGATGTGAAAGAGAAATATC  
 4051 AGAGAGTCAGAAATGACAAAGAGGATGATTTCACAATACCTAGAAAACAT  
 4101 GTAACCTATTTCAAACAGTCCTAAAATCAGAGCATTGAGATCAGACATAT  
 4151 CCTAATTAATGCTGTTGAAATAAATCACGTTGGGAAAACCTTTAACAATAT  
 4201 CTAAATTATCCCTAGGGTCAATTCACAGGAACATTCTCAAATCCCAAAC  
 4251 CGCAAAATAA

## MBHB-Rezept-7: SERALPHA Aminosäuresequenz

	10	20	30	40	50	60
1	MTSTCTNSTR	ESNSSHTCMP	LSKMPISLAH	GIIRSTVLVI	FLAASFVGNL	VLALVLQRKP
61	QLLQVTNRFI	FNLLVTDLLQ	ISLVAPWVVA	TSVPLFWPLN	SHFCTALVSL	THLFAFASVN
121	TIVVVSVDRY	LSIIHPLSYP	SKMTQRRGYL	LLYGTWIVAI	LQSTPPLYGW	GQAADFDERNA
181	LCSMIWGASP	SYTILSVVSF	IVIPLIVMIA	CYSVVFCAAR	RQHALLYNVK	RHSLEVRVKD
241	CVENEDEEGA	EKKEEFQDES	EFRRQHEGEV	KAKEGRMEAK	DGSLKAKEGS	TGTSESSVEA
301	RGSEEVRESS	TVASDGSMEG	KEGSTKVEEN	SMKADKGRTE	VNQCSIDLGE	DDMEFGEDDI
361	NFSEDDVEAV	NIPESLPPSR	RNSNSNPPLP	RCYQCKAAKV	IFIIIFSIVL	SLGPYCFLAV
421	LAVWVDVETQ	VPQWVITIII	WLFFLQCCIH	PYVYGYMHKT	IKKEIQDMLK	KFFCKEKPPK
481	EDSHPDLPPT	EGGTGKIVP	SYDSATFP			

sequence.seq is a single sequence

Using hydrophobicity file: /home/servers/server/toppredm/lib/Engelman-Steitz.sc

Using cyt/ext file: /home/servers/server/toppredm/lib/Cyt-Ext.prok

Using sequence file: sequence.seq

5

(1 sequences)

MTSTCTNSTRESNSSHTCMPLSKMPISLAHGIIRSTVLVIFLAASFVGN  
VLALVLQRKPQLLQVTNRFIENLLVTDLLQISLVAPWVATSVPLFWPLN  
SHFCTALVSLTHLFAFASVNTIVVSVDRYLSIIHPLSYPSKMTQRRGYL  
LLYGTWIVAILQSTPPLYGWGQAADFERNALCSMIWGASPSYILSVVSF  
IVIPLIVMIACYSVVFCARRQHALLYNVVRHSLEVRVKDCVENEDEEGA  
EKKEEFQDESEFRQHEGEVKAKEGRMEAKDGSLLKAKEGSTGTSESSVEA  
RGSEEVRESSTVASDGSMEGKEGSTKVEENSMKADKGRTEVNQCSIDLGE  
DDMEFGEDDINFSEDDVEAVNIPELPPSRNRNSNSNPPLPRCYQCKAAKV  
IFIIIFSIVLSLGPYCFLAVLAVWVDVETQVPQWVITIIWLFFLQCCIH  
PYVGYGMHKTIKKEIQDMLKKFFCKEKKPKEDSHPDLPGETEGGTGKIVP  
SYDSATFP

10

15

(p)roaryotic or (e)ukaryotic: p

20

Charge-pair energy: 0

Length of full window (odd number!): 21

Length of core window (odd number!): 11

25

Number of residues to add to each end of helix: 1

Critical length: 60

Upper cutoff for candidates: 1

30

Lower cutoff for candidates: 0.6

Total of 1 structures are to be tested

Candidate membrane-spanning segments:

35

Helix	Begin	End	Score	Certainty
1	36	56	1.964	Certain
2	79	99	1.652	Certain
3	102	122	1.564	Certain
4	148	168	1.264	Certain
5	193	213	2.320	Certain
6	400	420	1.879	Certain
7	431	451	1.854	Certain

40

-----  
Structure 1

45

Transmembrane segments included in this structure:

Segment	1	2	3	4	5	6	7
Loop length	35	22	2	25	24	186	10
K+R profile	4.00	0.00	1.00	0.00	9.00		

50

CYT-EXT prof

For CYT-EXT profile neg. values indicate cytoplasmic preference.

K+R difference: -11.00

55

Tm probability: 1.00

-> Orientation: N-out

## Patentansprüche

60

1. Das dargestellte Gen inclusive des 5' und 3' nichttranslatierten Bereichs
2. Transkriptionsfaktoren, RNA Polymerasen und Pharmaka sowie Chemikalien die die Expression des Gens in positiver oder negativer Weise beeinflussen.
3. Die von dem Gen transkribierte messenger RNA inclusive davon abgeleitete Spleißvarianten und Isoformen.
4. Die von der mRNA oder von dem intronlosen Gen abgeleitete cDNA.
5. Das von der mRNA (oder dem Gen oder der cDNA) abgeleitete oder hergestellte Protein.
6. Antikörper oder Antiseren welche gegen einzelne oder mehrere Epitope des Proteins oder gegen das ganze Pro-

65

tein hergestellt werden.

7. Monoklonale Antikörper oder Antiseren die gegen einzelne oder mehrere Epitope des Proteins oder gegen das ganze Protein hergestellt werden.

8. Expressionssysteme (eukaryotische Zelllinien, Hefezellen, Xenopus Oocyten, Baculovirussysteme, bakterielle Expressionssysteme) welche das genannte Protein exprimieren (nativ oder recombinant)

9. Ligand Bindungsstudien und Screening assays an nativen oder recombinanten Rezeptoren oder Zellen oder Zellmembranen welche diesen Rezeptor enthalten.

10. Transgene Tiere und knock out Tiere welche diesen Rezeptor in veränderter Dichte oder gar nicht exprimieren.

11. Methoden der Gentherapie welche sich auf diesen Rezeptor bzw sein Gen oder seine mRNA (cDNA) erstrecken und deren Entwicklung und Anwendung.

12. Sense und Antisense Oligonukleotide welche von diesem Gen abgeleitet wurden sowie deren Anwendung.

13. Die Diagnose und Behandlung von Krankheiten die mit diesem Rezeptor in direkter oder indirekter Weise in Verbindung stehen.

14. Methoden zur Diagnose von Erkrankungen die mit diesem Rezeptor (oder dessen Gen, mRNA) in direkter oder indirekter Weise in Verbindung stehen wie z. B. Hybridisierungstechniken, Sequenzierung, SSCP, RFLP, Northern Blot, Southern Blot, Western Blot, Expressions Arrays, Antikörper, Mutationsanalysen.

15. Die Benutzung der Informationen, oder des Rezeptors oder Zellen welche diesen exprimieren zur Entwicklung neuer Pharmaka, Verbindungen und Chemikalien und die Evaluierung vorhandener Pharmaka, Verbindungen und Chemikalien sowie zur Entwicklung neuer Technologien oder Evaluierung vorhandener Technologien.

16. Das Patent soll sich auch erstrecken auf die Punkte 1. bis 15. für modifizierte Proteine, und Gen, cDNA und mRNA Sequenzen (Aminosäureaustausche, Basenaustausche).

---

Hierzu 1 Seite(n) Zeichnungen

---

- Leerseite -

